How can the Data Revolution contribute to climate action in smallholder agriculture?

Jacob van Etten, Jonathan Steinke, and Mark van Wijk

Dr Jacob van Etten is leader of the Information Services and Seed Supplies component of Bioversity International. He holds a PhD from Wageningen University and previously worked at FAO, IRRI and IE University (Spain). He conducts research on digital media to support knowledge generation on agrobiodiversity-based technologies. j.vanetten@cgiar.org

Jonathan Steinke is a research fellow with Bioversity International and PhD candidate at Humboldt-University Berlin and the Leibniz Centre for Agricultural Landscape Research. His research is dedicated to the development of information and communication technology (ICT) and data-driven tools for effective agricultural advisory services. j.steinke@cgiar.org

Dr Mark van Wijk is senior scientist in Farming Systems Analyses at the International Livestock Research Institute. He holds a PhD from the University of Amsterdam, and previously worked at Wageningen University. In his work he analyses smallholder farm livelihoods across the developing world, with a special focus on quantifying the relationships between agricultural production options and indicators of human welfare. M.VanWijk@cgiar.org

Abstract

In this article, we discuss the ongoing Data Revolution in relation to climate action in agriculture. Data are highly relevant for climate action, as climate change makes current local knowledge increasingly irrelevant and requires smarter management of agricultural systems. We discuss five data-related concepts and explore how they are linked with agricultural climate action: lean data, crowdsourcing, big data, ubiquitous computing, and information design. We show practical examples for each of these concepts. There are many opportunities for improving agricultural development projects, providing new services to smallholder farmers, and generating better information for policy- and decision-making. Making the Data Revolution work for smallholder farmers' climate action not only takes further technological development, but also requires careful governance and public investment to avoid a few actors taking over the current innovation space and stifle further development.

Introduction

Climate change is among the most important challenges facing humanity today. Fortunately, climate change emerges at the same time as another major change in human history. This is the Data Revolution, epitomised by the unprecedented amounts of data produced by a wide range of means: satellite sensors, GPS, social media, wearable devices, and so forth. It was the Industrial Revolution that started the massive use of fossil energy that generated the climate problem. Perhaps the Data Revolution can provide the solution?

Data are already key for renewable energy management. Varying levels of sun and wind need to be carefully matched to the fluctuating energy needs of users. So-called 'smart energy grids' feed on data. They constantly measure and forecast energy production and energy use patterns, match supply and demand, and detect energy waste. Also in the agricultural sector, climatic information has become an important focus of innovation in modern systems. In 2013, *Monsanto* took over the *Climate Corporation*, a United States-based climate information service provider, for US$ 1 billion. This made the strategic value of climate data highly evident.

But what about smallholder farming in the (sub-) tropics? Climate change destroys information: local knowledge gradually loses its value as rainfall patterns change and new pests and diseases appear (Quiggin & Horowitz, 2003). The creation of new knowledge, adaptive management, and 'smart' management will require constant data flows. Obviously, the digital divide between rich and poor affects what is possible. Even so, the worldwide, steady expansion of mobile networks is making digital communication more and more accessible to smallholder farmers. Mobile networks have made rural communities leapfrog directly to mobile banking in certain countries. What is needed to make a similar shift in agriculture?

We believe that five emerging concepts related to the data revolution are key in this context. In the following sections, we explore these concepts to understand ongoing efforts and the future potential of data-driven approaches to agricultural climate action in smallholder agriculture. Although we list these concepts here separately, the description of each of the concepts will make clear that they are highly interconnected.

Key concept 1: Lean data

The idea of *lean data* emerged to address the need to monitor the social and environmental impacts of investments. Often, efforts in these areas are evaluated when they are well underway. This limits the degree of learning during their implementation and the scope of adjustments that can be made. Lean data involves using digital means to collect a minimalistic set of indicators at a frequent rate that allow monitoring of what is going on. For example, *constituent voice* measurements use very simple means to retrieve information about the perception of key stakeholders in change processes. Using simple Likert scales, participants indicate how they feel about the intervention in which they are involved. This allows project managers to keep their finger on the pulse. If they observe sudden changes or trends in the data, they can further investigate the causes through more qualitative inquiry.

Another interesting lean data idea has been piloted by the International Centre for Tropical Agriculture (CIAT). The *5Q* concept serves 'real-time' project monitoring using mobile telephone surveys, collecting the feedback of beneficiaries (Figure 1). Farmers respond via mobile phone to ultra-short questionnaires that are administered through automated voice response. By making questions conditional on the answers to previous questions, rich information can be obtained even though each farmer only answers five questions at a time (hence the name *5Q*). The information can be used for timely corrective action during the project cycle. The pilot found some limitations in the ability to synchronise the survey with ongoing field activities, but showed the potential of the *5Q* approach (Jarvis *et al*, 2015).

Figure 1. A local extension agent conducts a *5Q* survey using a tablet. (Photo: Manon Koningstein, CIAT)

The lean data approach has been applied more specifically to agriculture in the *Rural Household Multiple Indicator Survey (RHoMIS)*, developed by a collaborative group from the CGIAR (Hammond *et al*, 2016). This survey format stems from the realisation that a small number of variables can predict household food security status (Frelat *et al*, 2016), and that similar sparse indicators are needed for other aspects of farm performance. Climate-smart agriculture (or sustainable agricultural intensification) is about managing the trade-offs (positive and negative) across a large number of indicators, trying to avoid progress on one indicator causing a negative impact on another. The multiple aspects that need to be managed include productivity, poverty, greenhouse gas emissions, food security, gender and social inclusion. Systems approaches are widely advocated to deal with the multidimensionality of climate change, but there is a need for easy-to-use, quantitative tools to underpin these approaches.

The *RHoMIS* selects standard indicators for each of these aspects that are validated and require a relatively small number of questions. By including only questions that contribute to calculating these indicators, it avoids the 'design by committee' syndrome, which often leads to long questionnaires that satisfy the curiosity of the experts but that are neither complete nor parsimonious. The use of standard indicators and digital data collection tools (Open Data Kit with Android devices) also makes it easy to process the data automatically. This makes the results of the data analysis directly available, enabling the use of the resulting insights immediately to target project interventions. The strength of *RHoMIS* is not that it provides in-depth insights into any of these aspects, but that it allows the broad study of relationships between the different aspects. More in-depth, focused studies could follow-up on particular aspects identified from an exploratory analysis of *RHoMIS* data. The *RHoMIS* is already widely used for target-setting, monitoring and evaluation purposes (Hammond *et al*, 2017). It shows how the lean data approach can not only support adaptive management, but can be combined with a systems approach, which is another strong need for climate action.

Key concept 2: Crowdsourcing

Crowdsourcing is another important concept in the ongoing Data Revolution. Many information-related tasks are still best done by people. Digital means, however, make it easy to distribute tasks to large groups of people, and retrieve and combine the results. This makes it possible to scale the realisation of information-based tasks to levels that were not possible before.

An example of one such task is the transcription of weather data. Many old records with handwritten meteorological records exist; ships would typically keep detailed records of weather conditions. These data are now of much value in calibrating climate models. However, text recognition software has trouble recognising handwriting from several centuries ago. The *Old Weather* project therefore employed online volunteers to transcribe these records (www.oldweather.org). People are motivated to contribute as volunteers for various reasons: the scientific value of the tasks or the personal connection to the persons who wrote these records. The project website makes the task part of a game-like challenge, in which records of achievement are being kept and people receive badges or roles depending on their contribution. Until now, 20,000 people have participated in transcribing millions of records. These data are added to the *International Comprehensive Ocean-Atmosphere Data Set (ICOADS)* climate database (Freeman *et al*, 2016).

The crowdsourcing idea has also been applied to experimentation for climate adaptation in the agricultural sciences. Bioversity International (CGIAR) developed the *triadic comparisons of technologies (tricot)* methodology to make it possible for large numbers of farmers to 'massively test' different technologies (van Etten, 2011; van Etten *et al*, 2017). In *tricot*, each farmer receives a combination of three technologies (for example, crop varieties or types of inputs); they then test and compare the technologies using a very simple on-farm trial format. By giving farmers different, partially overlapping combinations of technologies, larger sets of technologies can be compared: for example, sets of 10-20 crop varieties. Crowdsourced field-testing not only expands the number of trials but also makes clever use of the diverse growing conditions of each field (in terms of weather, soil, planting date, other management choices) to analyse environmental adaptation in a single year. Crowdsourcing provides a bottom-up, data-intensive approach to climate adaptation, which should complement more top-down approaches, based on causal modelling. The strength of the crowdsourcing *tricot* approach lies in its external validity. Crop models are calibrated with data produced on agricultural research stations, which may not represent real conditions on farmers' fields. In contrast, the *tricot* approach samples a wide range of farm conditions that actually occur locally.

Key concept 3: Big data

The term *big data* denotes the massive quantity of data that are produced by humans interacting with digital media, by sensors, by business transactions, crowdsourcing, gene sequencing, *etc*. There are different definitions of big data around and there is overlap with the other key concepts in this paper, but the term big data emphasises the data management challenges that this data deluge has caused, as well as the emerging possibilities. For example, opportunities arise from data that are being generated as a side-product of other processes. Examples include digital transactions (online purchases, mobile money transfer, credit card use, *etc*), the clicking behaviour of website visitors, the terms used by search machine users, messages shared through social media, loyalty card use, and so on. Big data also results from the digitisation of data that were previously only available in analogue format (texts, images, audiovisual materials) or by adding a common structure to data that consists of separate small datasets. Big data generates many opportunities for innovative data analysis, for example by combining data from different sources or by repurposing data to detect real-time trends in time and space.

Big data is different from scientific data. Big data tends to rely on less control over sampling or observation. But the wide coverage or real-time nature of big data may override concerns about representativeness or the lack of experimental control. For example, social media users may not be representative of the world population, but they constitute such a large group that the data they produce may be relevant even if not fully representative. Science was traditionally based on deriving conclusions from scarce data through model-driven inference. Now, new methods are needed to deal with big data. At the same time, the limitations and risks of using big data need to be taken into account and better studied. Due to its limitations in terms of representativeness as well as ownership and privacy issues, big data will not completely substitute 'small data' studies but rather complement these (Kitchin, 2016).

For climate action in agriculture, it is clear that big data approaches have promise. For example, Simko & Pechenick (2010) present a method to aggregate crop trial data from different crop breeding trials, in spite of differences in experimental conditions, rating scales or proxies used. Lobell *et al* (2011) have shown that existing crop trial data can be repurposed to study the effects of climate on crop yield. Different efforts are underway to create consistent databases with crop trial data, standardising data formats. Data standardisation requires the development of 'ontologies', which are documented standards that describe the underlying elements and variables that are contained in the datasets and how these different elements/variables are interlinked (Shrestha *et al*, 2012). Big data approaches are still incipient for applications in smallholder agriculture and well-coordinated efforts are needed to achieve their full potential.

One important product that shows the power of big data for agriculture is the *Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)* dataset (Funk *et al*, 2014). This dataset is based on the interpolation of precipitation data produced by weather stations combined with satellite radar data and goes back several decades. The resulting dataset is quasi-global and gives daily precipitation estimates on a 5 x 5 km resolution grid. An important achievement is to source data from national meteorological institutes and others sources, which requires an important investment in social capital, as the availability of public meteorological data under unrestrictive licenses is on the decrease (Ramirez, 2012).

Key concept 4: Ubiquitous computing

The idea of *ubiquitous computing* is the opposite of the usual practice of concentrating computing mainly in a single device (PC or laptop), and shaping our tasks around this technology. Instead, it proposes to embed computing directly into use objects to integrate the digital devices into the routines of users. The idea of ubiquitous computing is closely related to (but not synonymous with) a number of other concepts, such as the *Internet of Things* (*eg* thermostats and light sensors talking with the lights and curtains in your house) and wearable devices (fitness watches, computing integrated into clothing and so on).

The idea of ubiquitous computing is interesting in smallholder farming because currently farmers often find it difficult to combine computing tasks with their daily practice. Important obstacles are illiteracy and the difficulty of finding a specific time and space for computing tasks. If data acquisition, processing, and feedback are fully integrated into the tools and tasks of farmers and designed according to their abilities and needs, it will be more likely that computing will positively affect their farming practice.

In modern farming, ubiquitous computing is already highly developed. Precision farming technologies make tractors constantly send and receive data to adjust planting density, fertilisation rates and so on within fields. Precision farming is an important part of climate action. Controlling input dosage, for example, can reduce wastage and avoid greenhouse gas emissions from fertilisers.

For smallholder agriculture, ubiquitous computing also holds promise, but is still in its infancy. One example is the development of the *Trans-African Hydro-Meteorological Observatory (TAHMO)* weather station network, which addresses the dearth of weather data in Africa. This initiative has designed a weather station that is extremely low in maintenance by avoiding any sensors with moving parts (Figure 2). It is connected to the mobile network to send the data it collects and is powered by a solar panel. These features help to overcome some of the main limitations of weather station networks in poor rural areas.

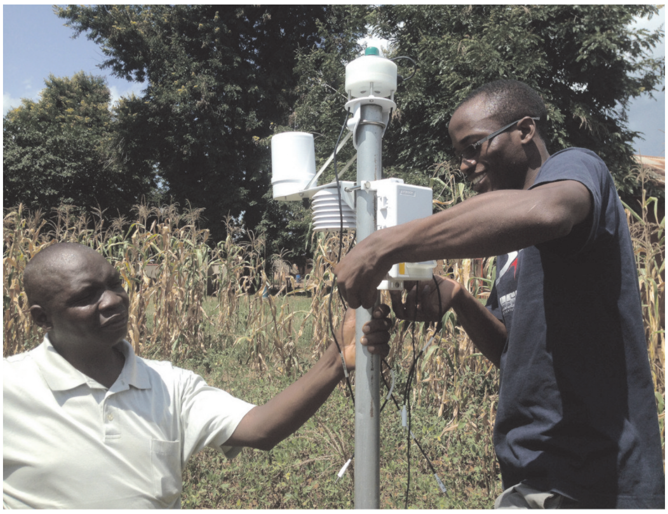

Figure 2. Installation of a meteorological station in Uganda by *TAHMO* engineer George Sserwadda. (Photo: *TAHMO*)

Another ubiquitous computing initiative, *Talking Plants*, focuses on practices around crop seeds, exploring the use of sensors to measure humidity of seeds in storage, and video as a medium to record farmers' information about their seeds, including their personal story, which often has much significance (Heitlinger *et al,* 2014).

It is evident from these two examples that careful design is needed for successful ubiquitous computing solutions. The design process needs to centre on the eventual users, taking into account their specific needs and interests, which may differ among users in terms of gender, age and other social factors. We believe that such design efforts would be very important in helping to bridge the current digital divide.

Key concept 5: Information design

Eventually, climate action depends on human decision-making, so it is crucial that data are converted into understandable information through *information design*. This concept refers to making data available in formats that allow users to derive insights to inform decisions. Over the last decade, complex, interactive visualisations have become available for personal computers, stimulating creativity to generate new visualisation formats. At the same time, scientists have made much progress in understanding how human visual perception works (Ware, 2013). Human visual perception is a pattern-seeking system that is intricately linked with human cognition. Interactive visualisation is being increasingly recognised as having a place in scientific knowledge generation. It should afford the discovery of new information by exploring the data and drilling down to get more detail (Ware, 2013).

For example, Steed *et al* (2013) argue that knowledge discovery from climate simulation data calls for increased visualisation capacity. Simulation data generates many models and scenarios, each producing output in the form of multiple variables. Data reduction as a preparation to then create simple visualisations can remove many of the features, precluding the generation of new insights. Steed *et al* (2013) created an analysis tool *(EDEN)* that includes interactive, multi-dimensional visualisation techniques that are more appropriate for the big data era.

We are not aware of parallel efforts in agricultural climate science that are at an advanced stage, however, the issues are very similar. We believe that more investment is needed in information design in agricultural climate studies.

Final remarks

It is clear that the Data Revolution is already underway to support climate action in smallholder agriculture. Many solutions are within reach from a technological perspective, but still require substantial effort and creativity to be adapted to smallholder agriculture through user-centred design. This involves building systems that respond to local problems with intensive feedback from future users; making the institutional arrangements, or generating business models, to make their use sustainable; and influencing the enabling environment so that these approaches gain long-term policy support and are embedded in solid regulations.

We think that data-intensive approaches are attractive for development investment. They can create practical solutions in agricultural climate action with concrete, visible benefits for farmers. Also, they generate business opportunities, create space for community initiatives, and provide entry points for more responsive policy. In other words, data-driven climate action provides opportunities for a wide range of actors, which could guarantee broad institutional support through an appeal to different institutional styles. From a climate action perspective, this broad appeal is a crucial success factor (Verweij *et al,* 2006).

In terms of policy and institutions, data governance is key, in order to balance privacy and data property rights with wider innovation possibilities provided by access to data. Innovation opportunities would quickly narrow if a few monopolistic players occupy the innovation space provided by the Data Revolution. Proactive policy, as well as public support and investment, will therefore be crucial in establishing an open space for business and community organisations in a way that will give rise to the interdependent, decentralised data management systems that are needed for agricultural climate action.

References

Freeman E, Woodruff SD, Worley SJ, Lubker SJ, Kent EC, Angel WE, Berry DI, Brohan P, Eastman R, Gates L, Gloeden W, Ji Z, Lawrimore J, Rayner NA, Rosenhagen G, Smith SR, 2016. ICOADS Release 3.0: a major update to the historical marine climate record. *International Journal of Climatology.* [https://doi.org/10.1002/joc.4775].

Frelat R, Lopez-Ridaura S, Giller KE, Herrero M, Douxchamps S, Djurfeldt AA, Erenstein O, Henderson B, Kassie M, Paul BK, Rigolot C, 2016. Drivers of household food availability in sub-Saharan Africa based on big data from small farms. *Proceedings of the National Academy of Sciences,* **113**(2), 458-463.

Funk CC, Peterson PJ, Landsfeld MF, Pedreros DH, Verdin JP, Rowland JD, Romero, BE, Husak GJ, Michaelsen JC, Verdin AP, 2014. A quasi-global precipitation time series for drought monitoring: *US Geological Survey Data Series 832*. [http://dx.doi.org/10.3133/ds832].

Hammond J, Fraval S, van Etten J, Suchini JG, Mercado L, Pagella T, Frelat R, Lannerstad M, Douxchamps S, Teufel N, Valbuena D, 2017. The Rural Household Multi-Indicator Survey (RHoMIS) for rapid characterisation of households to inform climate smart agriculture interventions: Description and applications in East Africa and Central America. *Agricultural Systems,* **151**, 225-233.

Heitlinger S, Bryan-Kinns N, Jefferies J, 2014. The Talking Plants: an interactive system for grassroots urban food-growing communities. *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, 459-462.

Jarvis A, Eitzinger A, Koningstein M, Benjamin T, Howland F, Andrieu N, Twyman J, Corner-Dolloff C, 2015. *Less is more: the 5Q approach*. Scientific Report. International Centre for Tropical Agriculture (CIAT). Cali, Colombia. [http://dapa.ciat.cgiar.org/] Accessed 16 February 2017.

Kitchin R, 2013. Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography,* **3**(3), 262-267.

Lobell D, Banziger M, Magorokosho C, Vivek B, 2011. Nonlinear heat effects on African maize as evidenced by historical yield trials. *Nature Climate Change,* **1**(1), 42-45.

Quiggin JC, Horowitz JK, 2003. Costs of adjustment to climate change. *Australian Journal of Agricultural and Resource Economics,* **47**, 429-446.

Ramirez-Villegas J, 2012. *Looking for weather data?* [http://dapa.ciat.cgiar.org/looking-for-weather-data/]. Accessed 16 February 2017.

Simko I, Pechenick DA, 2010. Combining partially ranked data in plant breeding and biology: I. Rank aggregating methods. *Communications in Biometry and Crop Science,* **5**(1), 41-55.

Shrestha R, Matteis L, Skofic M, Portugal A, McLaren G, Hyman G, Arnaud E, 2012. Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice. *Frontiers in Physiology,* **3**(326), 2.

Steed CA, Ricciuto DM, Shipman G, Smith B, Thornton PE, Wang D, Shi X, Williams DN, 2013. Big data visual analytics for exploratory earth system simulation analysis. *Computers & Geosciences,* **61**, 71-82.

van Etten J, 2011. Crowdsourcing crop improvement in Sub-Saharan Africa: a proposal for a scalable and inclusive approach to food security. *IDS Bulletin*, **42**(4), 102-110.

van Etten J, Beza E, Calderer L, Duijvendijk K, Fadda C, Fantahun, B, Kidane YG, van de Gevel J, Gupta A, Mengistu DK, Kiambi D, Mathur PN, Mercado L, Mittra S, Mollel MJ, Rosas JC, Steinke J, Suchini JG, Zimmerer KS, 2017. First experiences with a novel farmer citizen science approach: crowdsourcing participatory variety selection through on-farm triadic comparisons of technologies (tricot). *Experimental Agriculture* [https://doi.org/10.1017/S0014479716000739].

Verweij M, Douglas M, Ellis R, Engel C, Hendriks F, Lohmann S, Ney S, Rayner S, Thompson M, 2006. Clumsy solution for a complex world: the case of climate change. *Public Administration,* **84**(4), 817-843.

Ware C, 2013. *Information visualization. Perception for design*. Morgan Kaufmann. Third Edition.

News from the Field

Adaptation for Smallholder Agriculture Programme (ASAP)

ASAP is the largest global programme focused on climate-smart agriculture (CSA) and adaptation to climate change by smallholder farmers, and it currently supports projects in 36 countries, see https://www.ifad.org/topic/asap/overview . The Programme is managed by the International Fund for Agricultural Development (IFAD) and has a budget of US$ 366 million. The United Kingdom's Department for International Development (DFID) is the largest contributor to *ASAP*, which is also supported by 10 other donors.

The overall goal of *ASAP* is to increase the resilience of poor smallholder farmers to climate change. It aims to achieve this by introducing, testing and scaling-up multi-benefit adaptation approaches, geared to the needs of farmers. The modality of *ASAP's* operation is to provide additional grant funding into the larger loan-based agriculture and rural development projects that are part of IFAD's regular portfolio developed in *IFAD 9* (2012-2014). *ASAP* grants are intended to make these projects 'climate-smart' so that the projects in their entirety, and their funding streams, contribute to the *ASAP* goal. *ASAP* is seen as a significant step to mainstreaming climate change in IFAD's entire portfolio, as witnessed in the emerging *IFAD 10* (2015-2018) portfolio.

Priority activities are agreed with host governments in the countries where *ASAP* is supporting projects. These include:

- Agricultural diversification strategies;
- Avoiding losses in production caused by climate-related pests and diseases;
- Rehabilitating and protecting soils from water stress and erosion;
- Protecting productive lands and rural infrastructure from extreme climate events;
- Improving management of green and blue water resources;
- Enhancing and diversifying access to clean energy sources;